UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/685,656	10/14/2003	W. Todd Daniell	030220; 190250-1300	5661

38823 7590 07/20/2010

AT&T Legal Department - TKHR
Attn: Patent Docketing
One AT&T Way
Room 2A-207
Bedminster, NJ 07921

EXAMINER	
MACILWINEN, JOHN MOORE JAIN	
ART UNIT	PAPER NUMBER
2442	

MAIL DATE	DELIVERY MODE
07/20/2010	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/685,656	DANIELL ET AL.	
	Examiner	Art Unit	
	JOHN MACILWINEN	2442	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *23 February 2010*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1,6,11-14,16,17 and 19-40* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,6,11-14,16,17 and 19-39* is/are rejected.

7)☒ Claim(s) *40* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *3/02/2010, 4/07/2010, 4/27/2010*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

Applicant's arguments filed 2/23/2010 have been fully considered and in view of the pending claim amendments are persuasive. However, after further consideration, a new grounds of rejection has been made in view of Chadwick (US 7,320,020 B2).

Allowable Subject Matter

Claim 40 is objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

Said claim elaborates on performing the character-by-character analysis discussed in claim 1. Though the prior art, particularly Files (An information retrieval system based on superimposed coding) does teach character-by-character analysis, as well as removing spaces in certain circumstances, making a determination as to whether or not a space is to be removed, based on said space being adjacent to a solitary “i” or “a” is neither discussed, suggested or anticipated by the prior art. Said limitation, discussed in claim 40, is evaluated in light of the entirety of claim 1. Prior art text parsing/tokenization schemes, as noted above, do discuss removing spaces under certain conditions, but as text parsing/tokenization schemes traditionally have little concern over things like formatting, or what words may or may not follow a character or word to be discarded/removed, the prior art does not place any emphasis on considering

that a space may or may not be adjacent to a single-character word such as ‘i' or ‘a’, as is recited in claim 40.

Claim Rejections - 35 USC § 101

35 U.S.C. 101 reads as follows:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

Claims 25 – 29 and 32 – 38 are rejected under 35 U.S.C. 101 because said claims appear to be directed to non-statutory subject matter. Said claims are directed to “computer-readable storage medium”. Applicant is requested to clarify that said medium is non-transitory, as Applicant’s current claim language can be considered to include transitory and thus non-statutory embodiments such as signals.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

1. Claims 24 and 31 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement.

2. Regarding claims 24 and 31, claim 24 recites element “means for receiving a first email” is a means (or step) plus function limitation that invokes 35 U.S.C. 112, sixth paragraph. However, the written description fails to clearly link or associate the disclosed structure, material, or acts to the claimed function

such that one of ordinary skill in the art would recognize what structure, material, or acts perform the claimed function.

Applicant's Specification, pg. 7 lines 22 - 23 recites that "an email application 185 is being loaded into memory … thereby permitting the workstation 176 to send and receive email". Page 8, lines 5 - 6 continue, to recite that a "network interface 190 provides the interface … to receive [and] to transmit…". Applicant thus discloses a variety of items that may be interpreted as the means for "receiving". It is unclear which precise items recited in Applicant's Specification are intended to represent said means.

Claim 24 also recites "means for searching", "means for removing", etc. The above logic may be applied to these and the remaining "means for" claim language of claim 24 and the "means for" claim language of claim 31.

Applicant is required to:

(a) Amend the claims so that the claim limitation will no longer be a means (or step) plus function limitation under 35 U.S.C. 112, sixth paragraph; or

(b) Amend the written description of the specification such that it clearly links or associates the corresponding structure, material, or acts to the claimed function without introducing any new matter (35 U.S.C. 132(a)); or

(c) State on the record where the corresponding structure, material, or acts are set forth in the written description of the specification that perform the claimed function. For more information, see 37 CFR 1.75(d) and MPEP §§ 608.01(o) and 2181.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

3. Claims 1 and 39 are rejected under 35 U.S.C. 103(a) as being unpatentable over Shipp (US 2004/0093384 A1) in view of in view of Milliken (US 2004/0073617 A1), Chadwick (US 7,320,020 B2), Sahami (A Bayesian Approach to Filtering Junk E-Mail), Woitaszek (Identifying Junk Electronic Mail in Microsoft Outlook with a Support Vector Machine), Devine (US 6,968,571 B2), Burdick (US 2004/0107189 A1), Burdick (US 2004/0107189 A1), Files (Files, J. and Huskey, H. An information retrieval system based on superimposed coding. AFIPS Joint Computer Conferences. Proceedings of the November 18-20, 1969, fall joint computer conference. 1969. pp. 423 - 431), Anderson (US 2004/0064537 A1), further view of Uuencode and MIME FAQ (http://web.archive.org/web/20021217052047/http://users.rcn.com/wussery/attach.html).

4. Regarding claim 1, Shipp shows a method comprising receiving a first email message from a simple mail transfer protocol (SMTP) server ([18-19])

the first email comprising:

a text body ([57-58, 61-73])

an SMTP email address that includes a user name and a domain name ([19-24])

tokenizing the text body to generate a plurality of body tokens representative of the words in the text body ([67-68, 87, 93, 113])

tokenizing the SMTP email address to generate an address token representative of the SMTP email address ([22, 39, 43, 63]);

tokenizing the domain name to generate a domain token that is representative of the domain name ([22, 39, 43, 63]);

generating a 128-bit MD5 hash ([93], where MD5 is inherently 128-bit)

categorizing the first email message as a function of the spam probability ([14])

and filtering a second email message ([127]).

Shipp does not explicitly show all of:

the first email message comprising displaying characters and non-displaying characters, the non-displaying characters including non-displaying comments and non-displaying control characters;

a 32-bit string;

an attachment;

searching for the non-displaying characters in the first email message;

removing the non-displaying characters, including the non-displaying comments and non-displaying control characters;

tokenizing the attachment to generate a token that is representative of the attachment;

generating a hash of the attachment;

determining a corresponding spam probability value for each of the

plurality of body tokens, the address token, and the attachment token;

determining whether at least one of the plurality of body tokens, the address token and the attachment token is present in a database of tokens and, in response to a determination that at least one of the plurality of body tokens, the address token, and the attachment token is present in the database of tokens;

updating the spam probability value of the plurality of body tokens, the address token, and the attachment token.

Milliken shows the first email message comprising displaying characters and non-displaying characters, the non-displaying characters including non-displaying comments and non-displaying control characters ([10-13, 51-53]);

a 32-bit string ([48]);

an attachment ([13, 68, 70]);

searching for the non-displaying characters in the first email message ([69]);

removing the non-displaying characters, including the non-displaying comments and non-displaying control characters ([69]);

tokenizing the attachment to generate a token that is representative of the attachment ([70]);

generating a hash of the attachment ([10-13, 52]);

determining a corresponding spam probability value for each of the plurality of body tokens, the address token, and the attachment token ([56, 59, 74]);

determining whether at least one of the plurality of body tokens, the address token and the attachment token is present in a database of tokens and, in response to a determination that at least one of the plurality of body tokens, the address token, and the attachment token is present in the database of tokens ([56, 59-60, 70]);

updating the spam probability value of the plurality of body tokens, the address token, and the attachment token ([56, 59-60, 70]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp with that of Milliken in order to utilize Milliken's more current guidance for handling message attachments in view of a more current understanding of the contents of spam.

Shipp in view of Milliken do not explicitly show all of: determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the plurality of tokens;

classifying the plurality of tokens as spam, non-spam or neutral;

selecting the predefined number of interesting tokens, to create selected interesting tokens, the selected interesting tokens being the plurality of tokens having a greatest non-neutral probability value

performing an analysis on the selected interesting tokens to generate a spam probability.

Chadwick shows determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the plurality of tokens (col. 5 lines 25 – 30, col. 8 lines 21 - 22);

classifying the plurality of tokens as spam, non-spam or neutral (col. 8 lines 10 – 20, col. 8 lines 50 – 60);

selecting the predefined number of interesting tokens, to create selected interesting tokens, the selected interesting tokens being the plurality of tokens having a greatest non-neutral probability value (col. 8 lines 10 - 22);

performing an analysis on the selected interesting tokens to generate a spam probability (Fig. 2A).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken with that of Chadwick in order to further improve the spam filtering process as well as to better conserve client resources (Chadwick, col. 2 lines 33 - 38).

Shipp in view of Milliken and Chadwick not show where the analysis is a Bayesian analysis.

Sahami shows where the analysis is a Bayesian analysis (pg. 2, col. 2; pg. 4, col. 2; pg. 6, col. 1).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken and Chadwick with that of Sahami in order to further improve the accuracy with which spam email is identified.

Shipp in view of Milliken, Chadwick and Sahami thus do show selecting a subset of the generated tokens based on probability value as well as where the interesting tokens are a subset of the generated tokens (Chadwick, col. 8 lines 10 – 20, col. 8 lines 50 – 60, Sahami, pg. 6, col. 1, paragraph 1), but do not show

explicitly show where the tokens are sorted in accordance with the corresponding determined spam probability value.

Woitaszek shows where the tokens are sorted in accordance with the corresponding determined spam probability value (Tables 4 and 5).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick and Sahami with that of Woitaszek in order to arrange the calculated values in a logical manner, enabling a simple method of extracting the most interesting results (Sahami's disclosure involving selecting said most interesting tokens) via simply taking the top occurring results in Woitaszek's sorted list, as well as to include the abilities to integrate the spam software into a commonly used email program (Woitaszek, Abstract, pg. 1 col. 2).

Shipp in view of Milliken, Chadwick, Sahami and Woitaszek do not show utilizing a 32-bit string in indicative of the length of the first email message.

Devine shows utilizing a 32-bit string in indicative of the length of the first email message. (col. 24 lines 52-67).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick, Sahami and Woitaszek in order to better identify message contents so as to facilitate leveraging common code for processing messages (Devine col. 23 lines 60-61).

Shipp in view of Milliken, Chadwick, Sahami, Woitaszek and Devine do show determining displaying characters (Milliken, [69]) but not show determining

non-alphabetic characters in the first email message,

filtering the determined non-alphabetic displaying characters from the first email message;

and generating a phonetic equivalent for each word that includes only alphabetic characters that have a phonetic equivalent.

Burdick shows determining non-alphabetic characters in the first email message ([14,98]),

filtering the determined non-alphabetic displaying characters from the first email message ([14,98]);

and generating a phonetic equivalent for each word that includes only alphabetic characters that have a phonetic equivalent ([14,56]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick, Sahami, Woitaszek and Devine with that of Burdick in order to ensure the data (that is, email message contents) is in good form before it is further processed, thus increasing the ease of using the data and its utility (Burdick, [2-4]).

Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine and of Burdick do not explicitly show all of a per-character analysis that recursively determines for each character whether a character is a non-alphabetic character,

if the character is a non-alphabetic character, whether the character is a space,

and if the character is a space, determine whether is space is adjacent to a solitary “i” or “a”.

Files shows a per-character analysis that recursively determines for each character whether a character is a non-alphabetic character (pg. 424, left column),

if the character is a non-alphabetic character, whether the character is a space (pg. 424, left column),

and if the character is a space, determine whether is space is adjacent to a solitary “i” or “a” (pg. 424, left column and Fig. 1, pg. 432).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine and Burdick with that of Files in order to more efficiently store and process information (Files, pg. 423).

Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick and Files do not show appending the 32-bit string (Devine, col. 24 lines 52 – 67) to the generated MD5 hash to produce a 160-bit number.

Anderson shows ([57-59]) appending an MD5 hash (inherently 128-bits) to network transmission size information (shown by Devine to be said 32-bit string, and where 32 +128 is inherently 160).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick and Files with that of Anderson in order to better uniquely identify messages (Anderson [57-09]), leading to improved message spam identification.

Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files and Anderson do not show UUencoding said 160-bit number to generate a token representative of the attachment.

Uuencode and MIME FAQ shows UUencoding a file.

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files and Anderson with that of Uuencode and MIME FAQ in order to store the message identification information (represented by the 160-bit number shown by Shipp in view of Devine, Milliken and Anderson) in a format easily exchanged over email (UUencode and MIME FAQ) since UUencoding produces an easily emailed file and since the disclosure of Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files and Anderson relates to email and files transferred over email. Furthermore, UUencoding is a prior art element, as shown in UUencode and MIME FAQ, and thus UUencoding the 160-bit number is combing a prior art element (UUencoding) to known methods (the known methods shown by Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files and Anderson) to yield predictable results (the results being a UUencoded item).

Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files, Anderson and Uuencode and MIME FAQ thus show all of claim 1.

5. Regarding claim 39, Shipp in view of Milliken, Chadwick, Sahami, Woitaszek, Devine, Burdick, Files, Anderson and Uuencode and MIME FAQ further show where the first email message is received at a computing device

(Shipp, [18-19]).

6. Claims 6, 16, 17, 23, 24 and 25 rejected under 35 U.S.C. 103(a) as being unpatentable over Shipp in view of Milliken, Chadwick and Woitaszek.

7. Regarding claim 6, Shipp shows a method comprising receiving, at a computing device, a first email message comprising a text body, an SMTP email address ([39,43,69]), and a domain name corresponding to the SMTP email address ([64,65]), the text body including displaying characters and non-displaying characters (Shipp, [57-58, 61-73])

tokenizing the SMTP email address to generate an address token representative of the displaying characters of the STMP email address ([39, 43,, 63])

tokenizing the domain name token to generate a domain token representative of the domain name ([22])

determining a corresponding spam probability value from the tokens ([14,76])

filtering a second email message ([127]).

Shipp does not explicitly show all of: searching for the non-displaying characters in the first email message;

removing the searched non-displaying characters, including the non-displaying comments and non-displaying control characters,

tokenizing the attachment to generate an attachment token that is representative of the attachment;

determining whether at least one of the tokens is present in a database of tokens, and in response to the determination that at least one of the tokens is present in the database of tokens,

updating the spam probability value of at least one of the tokens.

Milliken shows searching for the non-displaying characters in the first email message ([69]);

removing the searched non-displaying characters, including the non-displaying comments and non-displaying control characters ([69]),

tokenizing the attachment to generate an attachment token that is representative of the attachment ([70]);

determining whether at least one of the tokens is present in a database of tokens, and in response to the determination that at least one of the tokens is present in the database of tokens ([56, 59, 74]),

updating the spam probability value of at least one of the tokens ([56, 59, 74]).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp with that of Milliken in order to utilize Milliken's more current guidance for handling message attachments in view of a more current understanding of the contents of spam.

Shipp in view of Milliken do not explicitly show all of: determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the tokens.

Chadwick shows determining a predefined number of interesting tokens,

the predefined number of interesting tokens being a subset of the tokens (col. 5 lines 25 - 30, col. 8 lines 21 - 22).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken with that of Chadwick in order to further improve the spam filtering process as well as to better conserve client resources (Chadwick, col. 2 lines 33 - 38).

Shipp in view of Milliken and Chadwick not show sorting the tokens in accordance with the corresponding probability values.

Woitaszek shows sorting the tokens in accordance with the corresponding probability values (Tables 4 and 5).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken and Chadwick with that of Woitaszek in order to arrange the calculated values in a logical manner, enabling a simple method of extracting the most interesting results (Chadwick's disclosure involving selecting said most interesting tokens) via simply taking the top occurring results in Woitaszek's sorted list, as well as to include the abilities to integrate the spam software into a commonly used email program (Woitaszek, Abstract, pg. 1 col. 2).

Shipp in view of Milliken, Chadwick and Woitaszek thus show all of claim 6.

8. Regarding claim 16, Shipp in view of Milliken, Chadwick and Woitaszek further show receiving the first email message including a text body (Milliken, [68]).

9. Regarding claim 17, Shipp in view of Milliken, Chadwick and Woitaszek further show tokenizing the words in the text body to generate body tokens representative of the words in the text body (Milliken, [68, 74]).

10. Claim 23 contains limitations addressed in the above rejection of claim 6. Claim 23 stands rejected for the reasons given in claim 6.

11. Claim 24 contains limitations addressed in the above rejection of claim 6. Claim 24 stands rejected for the reasons given in claim 6.

12. Claim 25 contains limitations addressed in the above rejection of claim 6. Claim 25 stands rejected for the reasons given in claim 6.

13. Claims 11 - 14, 19 – 22 and 26 - 29 rejected under 35 U.S.C. 103(a) as being unpatentable over Shipp in view of Milliken, Chadwick and Woitaszek as applied to claim 6 above, and further in view of Sahami.

14. Regarding claim 11, Shipp in view of Milliken, Chadwick and Woitaszek shows assigning an address spam probability value to the address token representative of the SMTP email address (Shipp, [22, 39, 43, 63], and Milliken, [56, 59-60, 70]);

assigning a domain spam probability value to the domain token representative of the domain name (Shipp, [22, 39, 43, 63], and Milliken, [56, 59-60, 70]); and

generating a probability value using the address spam probability and the domain spam probability assigned to the address token and the domain token (Shipp, [120]).

Shipp in view of Milliken, Chadwick and Woitaszek do not explicitly show generating a Bayesian probability.

Sahami shows generating a Bayesian probability (pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick and Woitaszek with that of Sahami in order to further improve the accuracy with which spam email is identified.

15. Regarding claim 12, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show comparing the Bayesian (Sahami, pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1) probability value with a predefined threshold value (Shipp, [14, 101, 123]).

16. Regarding claim 13, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show categorizing the first email message as spam in response to the Bayesian (Sahami, pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1) probability value being greater than the predefined threshold (Shipp, [14, 101, 123]).

17. Regarding claim 14, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show categorizing the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold (Sahami, pg. 6 col. 1).

18. Regarding claim 19, Shipp in view of Milliken, Chadwick and Woitaszek show assigning a body spam probability value to each of the body tokens representative of the words in the text body;

assigning an attachment spam probability value to the attachment token representative of the attachment; and

generating a probability value using the body spam probability value and the attachment spam probability value assigned to the body tokens and the attachment token (Milliken, [56, 59-60, 70]).

Shipp in view of Milliken, Chadwick and Woitaszek do not explicitly show: generating a Bayesian probability.

Sahami shows generating a Bayesian probability (pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick and Woitaszek with that of Sahami in order to further improve the accuracy with which spam email is identified.

19. Regarding claim 20, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show comparing the Bayesian probability value with a predefined threshold value (Sahami, pg. 4 col. 2).

20. Regarding claim 21, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show categorizing the first email message as spam in response to the Bayesian probability value being greater than the predefined threshold (Sahami, pg. 2 col. 2, pg. 4 col. 2 , pg. 6 col. 1).

21. Regarding claim 22, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show categorizing the first email message as non-spam in

response to the Bayesian probability value being not greater than the predefined threshold (Sahami, pg. 6 col. 1).

22. Regarding claim 26, Shipp in view of Milliken, Chadwick and Woitaszek show

Shipp in view of Milliken, Chadwick and Woitaszek do not explicitly show:

Sahami shows

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Shipp in view of Milliken, Chadwick and Woitaszek with that of Sahami in order to further improve the accuracy with which spam email is identified.

23. Claim 26 contains limitations addressed in the above rejection of claim 11. Claim 26 stands rejected for the reasons given in claim 11.

24. Claim 27 contains limitations addressed in the above rejection of claim 12. Claim 27 stands rejected for the reasons given in claim 12.

25. Claim 28 contains limitations addressed in the above rejection of claim 13. Claim 28 stands rejected for the reasons given in claim 13.

26. Claim 29 contains limitations addressed in the above rejection of claim 14. Claim 29 stands rejected for the reasons given in claim 14.

27. Claims 30, 31, 32, 33 and 34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Milliken in view of Chadwick and Woitaszek.

28. Regarding claim 30, Milliken shows a system comprising a memory component that stores at least the following: email receiving logic configured to

receive a first email message comprising an attachment ([51-53, 70]) and an address, the email message further including displaying characters and non-displaying characters ([42, 50, 59-60, 69])

search logic configured to search for the non-displaying characters in the first email message ([69]);

remove logic configured to remove the non-displaying characters, including the non-displaying comments and non-displaying characters ([69]);

tokenize logic configured to generate at least one attachment token representative of the attachment ([70]);

analysis logic configured to determine a corresponding spam probability value from the at least one attachment token ([56,59,70,74]);

database determining logic configured to determine whether the at least one attachment token is present in a database of tokens, and, in response to a determination that the at least one attachment token is present in the database of tokens ([56, 59, 70, 74]);

update the corresponding spam probability value of the at least one attachment token ([56, 59, 70, 74]);

wherein only the displaying characters are tokenized ([69]);

and filter a second email message ([14, 26]).

Milliken does not explicitly show all of: determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the tokens.

Chadwick shows determining a predefined number of interesting tokens,

the predefined number of interesting tokens being a subset of the tokens (col. 5 lines 25 – 30, col. 8 lines 21 - 22).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Milliken with that of Chadwick in order to further improve the spam filtering process as well as to better conserve client resources (Chadwick, col. 2 lines 33 - 38).

Milliken in view of Chadwick thus do show selecting a subset of the generated tokens based on probability value as well as where the interesting tokens are a subset of the generated tokens (Chadwick, col. 8 lines 10 – 20, col. 8 lines 50 – 60), but do not show explicitly show where the tokens are sorted in accordance with the corresponding determined spam probability value.

Woitaszek shows where the tokens are sorted in accordance with the corresponding determined spam probability value (Tables 4 and 5).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Milliken in view of Chadwick with that of Woitaszek in order to arrange the calculated values in a logical manner, enabling a simple method of extracting the most interesting results (Chadwick's disclosure involving selecting said most interesting tokens) via simply taking the top occurring results in Woitaszek's sorted list, as well as to include the abilities to integrate the spam software into a commonly used email program (Woitaszek, Abstract, pg. 1 col. 2).

29. Regarding claim 31, Milliken shows means for receiving a first email message comprising an attachment and an address, the first email message

further including displaying characters and non-displaying characters ([42, 50, 59-60]);

means for searching for non-displaying characters in the first email message ([69]);

means for removing the non-displaying characters, including non-displaying comments and non-displaying control characters ([69]);

means for generating at least one attachment token representative of the attachment ([70]);

means for determining a spam probability value from the at least one attachment token ([56, 59, 70, 74]);

means for, in response to a determination that the at least one attachment token is present in a database of tokens ([56, 59, 70, 74]); updating the spam probability value of the at least one attachment token ([56, 59, 70, 74])

wherein only displaying characters are tokenized ([14, 26, 69]) and

filtering a second email message ([14, 26, 69]).

Milliken does not explicitly show all of: determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the generated tokens.

Chadwick shows determining a predefined number of interesting tokens, the predefined number of interesting tokens being a subset of the generated tokens (col. 5 lines 25 – 30, col. 8 lines 21 - 22).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Milliken with that of Chadwick in order to

further improve the spam filtering process as well as to better conserve client resources (Chadwick, col. 2 lines 33 - 38).

Milliken in view of Chadwick thus do show selecting a subset of the generated tokens based on probability value as well as where the interesting tokens are a subset of the generated tokens (Chadwick, col. 8 lines 10 – 20, col. 8 lines 50 – 60), but do not show explicitly show where the tokens are sorted in accordance with the corresponding determined spam probability value.

Woitaszek shows where the tokens are sorted in accordance with the corresponding determined spam probability value (Tables 4 and 5).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Milliken in view of Chadwick with that of Woitaszek in order to arrange the calculated values in a logical manner, enabling a simple method of extracting the most interesting results (Chadwick's disclosure involving selecting said most interesting tokens) via simply taking the top occurring results in Woitaszek's sorted list, as well as to include the abilities to integrate the spam software into a commonly used email program (Woitaszek, Abstract, pg. 1 col. 2).

30. Claim 31 contains limitations addressed in the above rejection of claim 31. Claim 32 stands rejected for the reasons given in claim 31.

31. Regarding claim 33, Milliken in view of Chadwick and Woitaszek further show receive the first email message having a text body (Milliken, [68]).

32. Regarding claim 34, Milliken in view of Chadwick and Woitaszek further show tokenize words in the text body to generate body tokens representative of

the words in the text body (Milliken, [68, 74]).

33. Claims 35 – 38 are rejected under 35 U.S.C. 103(a) as being unpatentable over Milliken in view of Chadwick and Woitaszek as applied to claim 32 above, and further in view of Sahami.

34. Regarding claim 35, Shipp in view of Milliken in view of Chadwick and Woitaszek shows assigning an address spam probability value to the address token representative of the SMTP email address (Milliken, [56, 59-60, 70], Chadwick, col. 6 lines 1 - 22);

assigning a domain spam probability value to the domain token representative of the domain name (Shipp, [22, 39, 43, 63], and Milliken, [56, 59-60, 70], Chadwick, col. 6 lines 1 - 22); and

generating a probability value using the address spam probability and the domain spam probability assigned to the address token and the domain token (Chadwick, col. 5 lines 25 - 30, col. 6 lines 1 – 22, Milliken, [60]),

Milliken in view of Chadwick and Woitaszek do not explicitly show generating a Bayesian probability.

Sahami shows generating a Bayesian probability (pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1).

It would have been obvious to one of ordinary skill in the art at the time of the invention to modify the disclosure of Milliken in view of Chadwick and Woitaszek with that of Sahami in order to further improve the accuracy with which spam email is identified.

35. Regarding claim 36, Milliken in view of Chadwick, Woitaszek and Sahami further show comparing the Bayesian (Sahami, pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1) probability value with a predefined threshold value (Chadwick, col. 3 lines 20 - 30).

36. Regarding claim 37, Milliken in view of Chadwick, Woitaszek and Sahami and Sahami further show categorizing the first email message as spam in response to the Bayesian (Sahami, pg. 2 col. 2, pg. 4 col. 2, pg. 6 col. 1) probability value being greater than the predefined threshold (Chadwick, col. 3 lines 20 - 30).

37. Regarding claim 38, Shipp in view of Milliken, Chadwick, Woitaszek and Sahami further show categorizing the first email message as non-spam in response to the Bayesian probability value being not greater than the predefined threshold (Sahami, pg. 6 col. 1).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to JOHN MACILWINEN whose telephone number is (571)272-9686. The examiner can normally be reached on M-F; 9:00-5:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Philip Lee can be reached on 571-272-3967. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

JOHN MACILWINEN

571-272-9686

/Philip C Lee/
Acting SPE of Art Unit 2442